JARQ 51 (1), 91 - 97 (2017) https://www.jircas.go.jp

Development of an Agricultural Field Study Database: For sharing multidisciplinary information on *in situ* photographs

Fukuhiro YAMASAKI[1], Makoto KAWASE[2] and Masaru TAKEYA[1]*

[1] Genetic Resources Center, National Agriculture and Food Research Organization (Tsukuba, Ibaraki 305-8602, Japan)
[2] Faculty of Life and Environmental Sciences, University of Tsukuba (Tsukuba, Ibaraki 305-8572, Japan)

Abstract
Through its website, the NARO Genebank (former NIAS Genebank) has been providing to the scientific community and general public different types of information regarding genetic resources including passport, characterization, evaluation and photographic data, in order to facilitate access to their conserved genetic resource records. However, the website currently lacks information on farming culture associated with the genetic resources at collection sites, as most of its currently accessible photographs are captured from experimental fields or laboratories. In contrast, photographs of *in situ* conditions taken in the fields of farmers during the exploration of plant genetic resources capture both biological and cultural information, such as crop morphology, geographical and habitat ecological conditions, farming implementations, ethnic costumes, people's lifestyles, cooking procedures, and other information. Unfortunately, most of these photographs are being hoarded and will likely never be accessible by the public. In order to disclose a large number of *in situ* photographs, we have developed a new web-retrievable database called the "Agricultural Field Study Database," in which photographs are searchable with keyword tags and computational image similarity that allow the user to browse and retrieve them, along with map data and relevant genetic resources information. This system will expand the information currently provided by the NARO Genebank with ethnobotanical information, thereby granting useful resources to a wider range of research fields, particularly disciplines related to anthropology and ethnobotany.

Discipline: Information technology
Additional key words: farming culture, genetic resources, web-based retrieval system

Introduction

The NARO Genebank, which was changed designation from the NIAS Genebank in April 2016, is the central coordinating project in Japan for the conservation of plants, microorganisms, animals, and DNA materials related to agriculture. It provides various information on genetic resources including passport, characterization, and evaluation data via its web retrieval database (Takeya et al. 2013a) to facilitate user access to conserved genetic resources. A large number of photographs are also available on its website, although most show close-ups of seeds and plants in a full length composition as taken in experiment fields or laboratories, while very few were taken showing *in situ* conditions in the fields of farmers, given the primary nationwide responsibility of the NARO Genebank to focus on *ex situ* conservation of genetic resources as breeding and research materials for agricultural development. However, it should be noted that *ex situ* conservation removes traditional cultivars from their original context (Jarvis 2000). On the other hand, this project has been dispatching several overseas exploration teams annually for collecting genetic resources (Okuno et al. 2005). Exploration for collecting plant materials from on-farm and *in situ* conditions is one of the important first steps for introducing genetic resources into gene banks. Agriculture not only produces

This work was supported by KAKENHI Grant Number 24405049.
*Corresponding author e-mail: katu@affrc.go.jp
Received 16 November 2015; accepted 10 March 2016.

food for human beings but also serves as the root of culture (Nakao 1966). Gene banks cannot conserve farming culture directly, but offer a general strategy for collecting in fields that entails lots of photographs (Hawkes et al. 2000). An *in situ* photograph contains plenty of information, such as the morphology of collected material, geographical and ecological information about the collection site, farming practices, ethnic costumes, people's lifestyles, and cooking procedures, and thus can be beneficial to ethnological study, not to mention agricultural research. We aim to contribute to a wider range of research fields by making those *in situ* photographs available through the Internet. A good but simple precedent is "Sasuke Nakao's Slide Database" on a website linked with the Library and Science Information Center of Osaka Prefecture University (Kojima et al. 2001) that publishes numerous photographs taken by Nakao, a famous Japanese botanist and ethnobotanist through his field trips from 1955 through 1984. The present study goes one step forward to design an "Agricultural Field Study Database" featuring a more user-friendly interface and providing links to relevant genetic resources and the Geographic Information System (GIS).

Materials and methods

1. Preparation of photographs

About 13,200 photographs taken in South and Southeast Asia during the exploration of genetic resources by Kawase, the second author, were used as the materials for developing the Agricultural Field Study Database. Given Kawase's participation in several plant exploration missions to collect crop landraces, observe cultivation practices, and note traditional utilization in the areas, those photographs are considered ideal materials for constructing this multidisciplinary database related to plant genetic resources. As film cameras were used from 1985 to 2006, we also had to digitize about 11,850 photographs as high resolution tiff files with a film scanner for protection against aging degradation. The digitized photographs and about 1,350 images taken recently with digital cameras were appended to the Adobe Lightroom catalog, along with attached keyword tags that represent the photographic subject(s) for enabling efficient search. The shooting date of each film was also recorded as a keyword tag with reference to a date stamp and field notes. Photographs were also exported to smaller web-optimized jpeg files after a quick correction.

2. Database implementation

A structured query language, MySQL was selected for the relational database management system in our new field study database because it is fast, multi-threaded, multi-user, and robust, and was adopted by the NARO Genebank for its web retrieval system (Takeya et al. 2013b).

(1) Schema

Figure 1 shows the current entity-relationship diagram of the Agricultural Field Study Database. Photo (independent entity) and tag (dependent entity) have a one-to-many relationship. The details of fs_photo_similarity will be discussed later.

(2) Extract data from jpeg files and insert into tables

Keyword tags are stored in the IPTC block (IPTC 2014). Using PHP with the GD library (http://php.net/manual/en/book.image.php) makes it possible to parse the binary IPTC block into single tags by code described in Fig. 2.

A jpeg file created by a digital camera usually contains various information in the Exif header (CIPA 2012). For example, when PHP code `$exif = exif_read_data($source, 'ANY_TAG');` runs, `$exif` will have the keys and values listed in Table 1. Latitude and longitude data are stored in degrees minutes seconds (DMS) format, in which decimal degrees are more suitable for treating GIS data; thus, we converted the latitude and longitude data into fractional values as follows: `1 * ((21/1) + (145479/10000)/60 + (0/1)/3600)`. It is important to note that when GPSLatitudeRef is "S" or

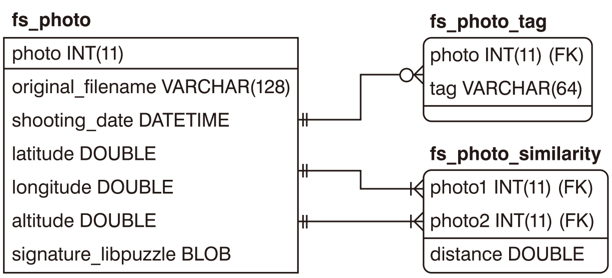

Fig. 1. Current entity-relationship diagram of the Agricultural Field Study Database.

```
$source = "example.jpg";
$size = getimagesize($source, $info);
if (isset($info['APP13'])) {
  $iptc = iptcparse($info['APP13']);
  if (isset($iptc['2#025'])) {
    foreach ($iptc['2#025'] as $tag) {
      //insert $tag into DB
    }
  }
}
```

Fig. 2. Example PHP code to parse the binary IPTC block into single tags.

GPSLongitudeRef is “W”, it must be multiplied by -1 to make it negative. We extracted keyword tags, latitude, longitude, altitude, and DateTimeOriginal data from all files, and then inserted the data into respective tables.

(3) Image similarity

To reinforce arbitrary classification by tagging, we introduced a computational image comparison algorithm—libpuzzle (http://www.pureftpd.org/project/libpuzzle)—based on the relative brightness of image regions (Wong et al. 2002) for computing the distance between all photograph pairs, and then inserted the data into table fs_photo_similarity. In almost all cases where a pair of photographs show a quite low distance value close to 0, both are shots of the same subject under the same conditions. It helps us to find duplicated photographs and reduce redundancy in the database. And as some photographs have a similar overall composition in a relatively small distance, their computational results will also prove useful for suggesting related photographs (Fig. 3).

3. Web-based retrieval system

(1) Top page

The top page shows ten thumbnails to make it easier for users to see what kinds of photographs are available. Samples will be chosen randomly, though the frequency of appearance is weighted by SQL as detailed in Fig. 4. The system tends to choose photographs that have major keyword tags.

Table 1. Sample output of the exif_read_data function

Key	Value
ExposureTime	1/125
FNumber	13/1
DateTimeOriginal	2014:07:12 09:26:50
GPSLatitudeRef	N
GPSLatitude	Array([0] => 21/1, [1] => 145479/10000, [2] => 0/1)
GPSLongitudeRef	E
GPSLongitude	Array([0] => 95/1, [1] => 8472/10000, [2] => 0/1)
GPSAltitude	7515/100

Distance: 0.10561578927066
Same subject, high similarity
(Both: Myanmar 2000)

Distance: 0.27077971952578
Different subject, higher similarity
(Up: India 1992, Down: Myanmar 2004)

Distance: 0.90451816904023
Different subject, low similarity
(Up: Pakistan 1991, Down: Myanmar 2005)

Fig. 3. Examples from the computational results of image similarity.

Major keyword tags are also shown at the bottom of the page as "Popular tags." Highly popular tags appear in bold type (Fig. 5). Users are able to move to a "details page" from sample photographs, and may also search for photographs by selecting a tag.

(2) Search results page

The search results page shows a list of thumbnails matching the search criteria. There are ten thumbnails per page, though the asynchronous JavaScript and XML (Ajax) pagination feature will help to see an arbitrary page with efficient data communication. Moreover, once a tag has been selected, a new list is displayed as "Related tags" instead of "Popular tags" (Fig. 6). This new list only displays tags related to the current selection and allows the search to be narrowed down. Already selected tags can be deselected by clicking on them.

```
    SELECT substring_index(
             group_concat(photo order by rand() separator '\t')
           , '\t', 1) as sample_photo
      FROM fs_photo_tag
  GROUP BY tag
  ORDER BY count(*) * rand() desc
     LIMIT 10
```

Fig. 4. SQL code for sample photograph selection.
Ten photographs will be chosen by weighted random sampling.

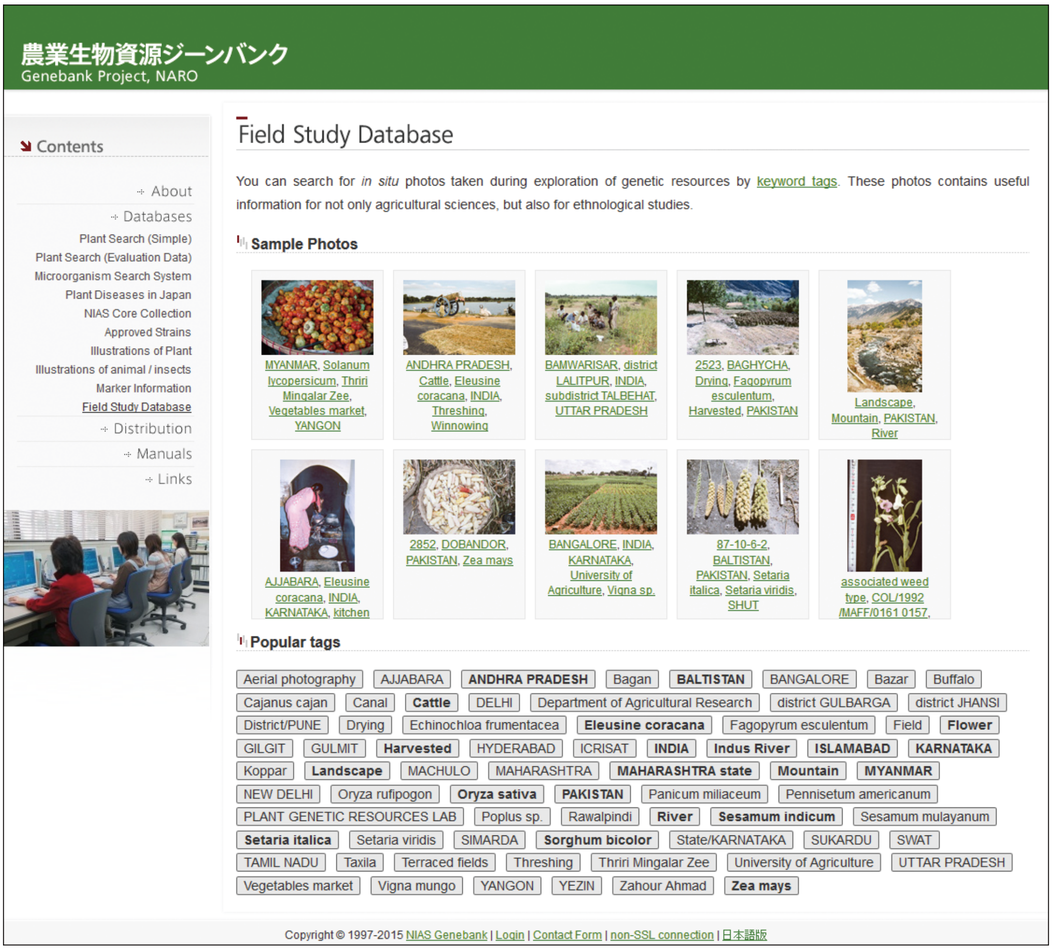

Fig. 5. Top page of the Agricultural Field Study Database.
Sample photographs and a list of popular tags are shown.

(3) Details page

This page shows a large-sized photograph with its additional information and related photographs. Additional information includes keyword tags, shooting date, shooting location, and relevant genetic resources. In cases where latitude and longitude data are available, Google maps appear inline. Related photographs are selected based on common keyword tags or low distance values (Fig. 7). Users can narrow down their search by adding a keyword tag, or may move to another details page from related photographs by clicking a thumbnail. When relevant genetic resources exist, a link is generated to detailed pages of the plant search system (http://www.gene.affrc.go.jp/?db_pl_e) that provide passport and evaluation data, and an online application form for distribution.

Results and discussion

We developed the web-accessible Agricultural Field Study Database in order to share the photographs taken during the exploration of plant genetic resources. The photographs are searchable with keyword tags, and users can obtain information about the shooting date, shooting location, and relevant genetic resources. Table 2 lists the major keyword tags and the number of currently published photographs related to those tags.

We will release and increase the number of photographs, while giving due consideration to portrait rights, copyrights, religious rules, and other issues. As this system features the efficient insertion of data, we are thus able to add many photographs with less labor. Kawase's already digitized field notes also contain useful information, and are expected to be published in a form that allows cross-referencing to photographs in the next version of the Agricultural Field Study Database. In terms of image similarity, there is room for improvement because the current version does not consider color information and only finds photographs having a similar overall composition. Adding the analysis of color histograms or a comparison of regions-of-interest could be an effective way to achieve improvement. Giving a high relevance rank to frequently viewed photographs using access analysis may also be useful. We strongly expect that this database could become a common platform for sharing information related to field studies by adding content and improving functions.

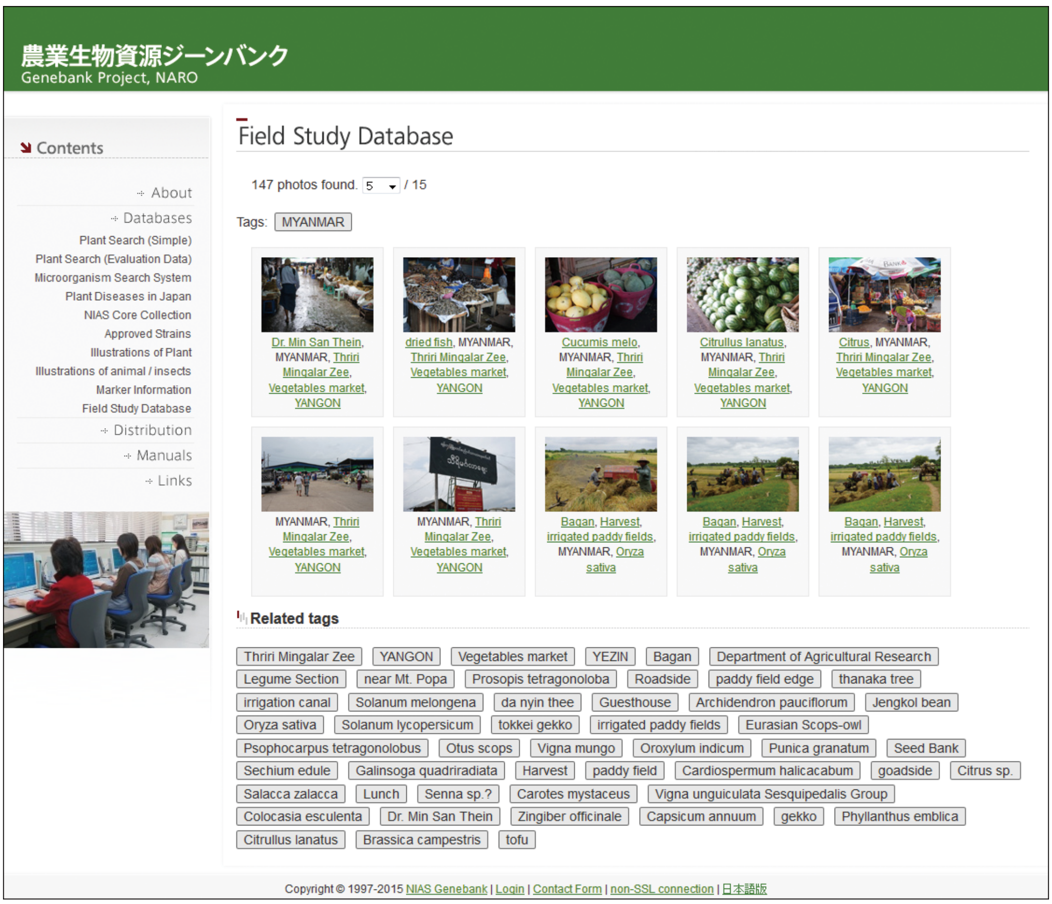

Fig. 6. Selected tags, matched photographs, and related tags are shown on the search results page.

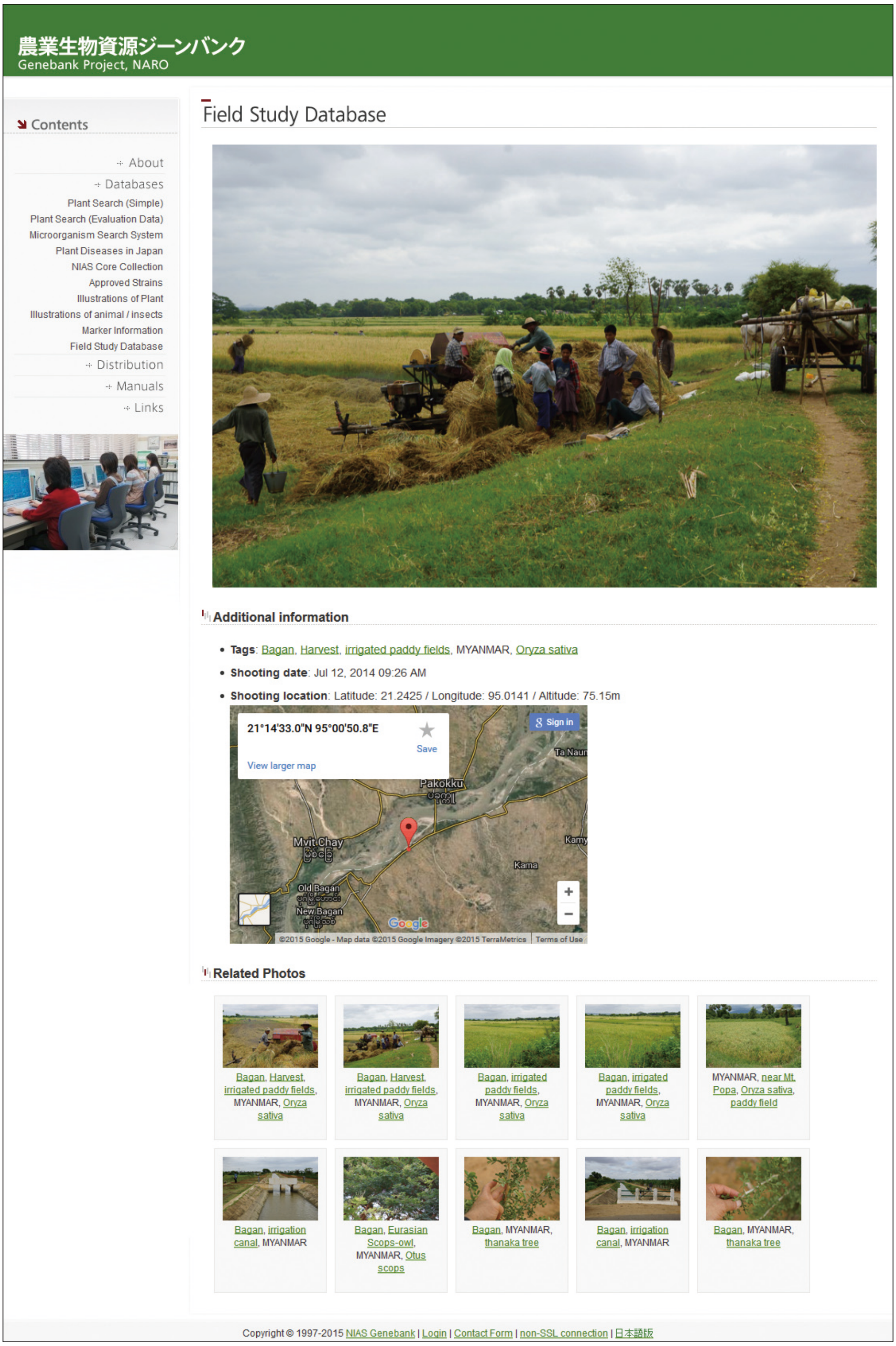

Fig. 7. Details page shows a large-sized photograph; additional information also includes inline Google maps (when latitude and longitude are available) and related photographs.

Table 2. Major tags and the number of related photographs

Tag	Number of photographs
PAKISTAN	2755
INDIA	2086
Landscape	1159
Mountain	824
KARNATAKA	506
Sesamum indicum	450
River	363
Setaria italica	355
Cattle	354
ANDHRA PRADESH	285
BALTISTAN	241
Oryza sativa	213
Zea mays	201

Acknowledgments

We would like to thank Ms. Natsumi Kishimoto for her patient work on digitization, and Dr. Ernesto Borrayo for his critical reading of the manuscript.

References

CIPA: Exchangeable image file format for digital still cameras: Exif Version 2.3 http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf

Hawkes, J. G. et al. (2000) *The Ex Situ Conservation of Plant Genetic Resources*. Kluwer Academic, Dordrecht, pp.250.

IPTC: IPTC Photo Metadata Standard. https://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata

Jarvis, D. I. et al. (2000) *A Training Guide for In Situ Conservation On-farm*. International Plant Genetic Resources Institute, Rome, pp.190.

Kojima, A. et al. (2001) Construction of Web-based Image Retrieval and Sharing System for Digital-Archive of Valuable Cultural Material. *IEICE technical report*, **101**(180), 61-68 [In Japanese].

Nakao, S. (1966) *Saibai shokubutsu to noukou no kigen*. Iwanami-shoten, Tokyo, pp.192 [In Japanese].

Okuno, K. et al. (2005) Plant genetic resources in Japan: platforms and destinations to conserve and utilize plant genetic diversity. *JARQ*, **39**, 231-237.

Takeya, M. et al. (2013a) Systems for making NIAS Core Collections, single-seed-derived germplasm, and plant photo images available to the research community. *Genetic Resources and Crop Evolution*, **60**(7), 1945-1951.

Takeya, M. et al. (2013b) Genebank data management software incorporating seed-viability test results. *Plant Genetic Resources*, **11**:217-220.

Wong, H. C. et al. (2002) An image signature for any kind of image. *Proc. of International Conference on Image Processing*, **2002**, 409-412.